

13 Contributions of Nonhuman Animal Models to Understanding Human Speech Perception

Keith R. Kluender
University of Wisconsin
Andrew J. Lotto
Boys Town National Research Hospital
Lori L. Holt
Carnegie-Mellon University

Broadly speaking, there are two ways nonhuman animal models contribute to our understanding of speech perception by humans—by analogy and by homology. The former is the generally easier task, and historical examples are more abundant. Because demonstrating strict homology requires deeper explication of underlying mechanisms, claims may be more precarious, but they carry greater explanatory potential.

When studying the nonhuman organism as analogy, emphasis is most often on how animal neurophysiological or behavioral processes have adapted to fulfill some requirement of a particular ecological niche. By contrast, study of nonhumans as homology must exceed the bounds of such niches in search of common underlying mechanisms across varying ecologies. In practice, this frequently involves undermining ecological integrity—presenting nonecological stimuli, intruding into the cranium, having subjects press bars or peck keys, and controlling experience in ways that are unethical with human participants.

When studying common underlying processes (homology), it also is true that nonhuman animals become a method more than an object of study. In service of revealing processes central to human speech perception, data from experiments with nonhuman subjects join a greater arsenal that also includes data from human perception studies and from computational simulations. The problem, not the organism, dictates these methods.

ABOUT SPECIALIZATIONS

Studies with nonhuman animals frequently have been directed toward putative specializations for perception of human speech (e.g., Liberman & Mattingly, 1989). In this vein, comparisons between

data from human and nonhuman subjects have been used both to buttress claims for human specialization (e.g., Liberman & Mattingly, 1985) and to challenge such claims (e.g., Kluender, Diehl, & Killeen, 1987; Kuhl & Miller, 1975, 1978).

The greatest evidence for human specialization for communication, however, can be found much more in mechanisms for speech production than in processes for speech perception (e.g., Fitch, 2000; Lieberman, 1984). Owing primarily to characteristics of supralaryngeal anatomy, the adult human possesses sound-producing abilities unrivaled among other organisms. This capacity is revealed in a grand assortment of speech sounds used contrastively by languages. The UCLA Phonological Segment Inventory Database (UPSID) contains a representative sample of the inventories of speech sounds used by the world's languages (Maddieson, 1984). Fifty-eight phonetic attributes are required to characterize all 558 consonants, 260 vowels, and 51 diphthongs in the UPSID sample. Such capacity dwarfs that of any other animal. The largest reported inventory of primate calls, by contrast, is 35 for the cotton-top tamarin (*Sanguinus oedipus*; Cleveland & Snowdon, 1982). Of course, the translation from nonhuman primate vocalizations to functional linguistic units used by humans is not obvious. However, it is safe to suggest that, owing to supralaryngeal anatomy, the number of functionally different mouth sounds for humans is more than an order of magnitude larger than for any other primate.

Whereas anatomical evidence suggests specializations supporting speech production by humans, whether humans are the genetic recipients of specialized processes for perception of human speech has been a good deal more contentious. In the 1970s, many researchers had specific ideas about what would constitute biological specializations for speech perception. The most significant example was Abbs and Sussman's (1971) suggestion that there may exist sensorineural configurations—feature detectors—that are specifically sensitive to properties of speech sounds that serve as trademarks of distinctive features. Eimas and Corbit (1973) were the first to propose an explicit feature detector model for speech perception. In large part, this early work and dozens of papers throughout the decade were inspired by analogy to electrophysiological findings with animals. Soon after the classic optic nerve studies of “bug perceivers” in *Rana pipiens* frogs (Lettvin, Maturana, McColloch, & Pitts, 1959) were studies of neurons in the auditory (VIIIth) nerve (Frishkopf & Goldstein, 1963) and thalamus (Mudry, 1978) of bullfrogs *Rana catesbeiana*. These units seemed ideally tuned to frequencies in bullfrog mating calls.

Only after about a decade of behavioral adaptation studies (which allegedly served to selectively fatigue feature detectors) with humans was it concluded by most investigators that there was little future in feature detectors and selective adaptation studies for revealing processes underlying speech perception. Experimental results proved problematic for feature detector models as proposed, and data were more readily explainable on general perceptual grounds. For some (e.g., Diehl, 1981; Diehl, Kluender, & Parker, 1985; Remez, 1979), the fall of feature detectors was public. For much of the field, however, research activity simply moved to more promising ideas and methods.

Despite the meager long-term impact of feature detectors for explaining processes of speech perception, some consideration of simpler biological systems has been sustained. Clear instances of specialized perceptual mechanisms for species-specific vocalizations have been amply demonstrated in a number of nonmammalian organisms including crickets (Hoy, Hahn, & Paul, 1977), cricket frogs (Ryan & Wilczynski, 1988), green tree frogs (Gerhardt & Rheinlander, 1982), zebra finches (Katz & Gurney, 1981), white-crowned sparrows (Margoliash, 1983), and canaries (Nottebohm, Stokes, & Leonard, 1976). As part of his arguments for a specialized speech module, Liberman (e.g., Liberman & Mattingly, 1985) appealed to several of these examples and to bats as apt analogies for specialized processes for human communication. By analogy to results from studies such as those noted earlier and particularly to bat biosonar, Suga (chap. 11, this volume) concludes that the human auditory system has developed highly specialized mechanisms for processing speech.

The simple fact that human languages use more than 800 different sounds argues against innate specializations for perceiving particular speech sounds. Although it is true that the situation

gets better if one describes the hundreds of speech sounds in terms of combinations drawn from a smaller group of features, even with the 58 used by Maddieson (1984) there is simply too much diversity in sounds used by languages to be plausibly accommodated by innate feature detectors. The typical language includes 20 to 37 consonants and vowels (about 31 on average), although a few get by with as few as 11. (One language, !Xu, has 141.) By contrast, innately specified and highly specialized mechanisms, as products of selective pressure, should give rise to substantial conformity in communication signals. Systems should adhere rather closely to a collection of signals that have been primed by the biological substrate. Furthermore, this collection should be relatively modest in size. After accommodating an inventory that is adequate for successful communication, there would be little pressure to increment the number of signals. These conditions are generally well met by nonhuman systems—analogs embraced by investigators such as Liberman and Suga. The problem, of course, is that human languages far exceed expectations from generally conservative processes of evolution. Whether cast as classical feature detectors or as more contemporary information-bearing elements (IBEs; Suga, chap. 11, this volume), highly specialized species-specific mechanisms revealed in studies of relatively simple organisms with very limited vocal repertoires (and stereotypic responses) must be viewed with suspicion, even as analogies to human vocal communication. This being said, although analogies to simple instinctual processes are not particularly revealing, other species-general processes—most notably perceptual learning—provide better comparisons between human and nonhuman performance.

AUDITORY DISCONTINUITIES AND CATEGORICAL PERCEPTION

Evidence suggesting that speech perception is not founded on neural predispositions for particular speech sounds does not, by itself, compel one to conclude that there exist no sensory predispositions serving to support communication. Short of proposing neural circuits such as feature detectors and IBEs (Suga, chap. 11, this volume), a relatively common broad claim is that languages take advantage of sensory discontinuities in the service of speech perception. The most well-known pattern of perceptual performance with speech sounds, categorical perception, has been seen as indicative of such sensory discontinuities (e.g., Kuhl, 1981; Kuhl & Miller, 1978).

Three features define categorical perception: a sharp labeling (identification) function, discontinuous discrimination performance (near perfect across identification boundary and near chance to either side), and the ability to predict discrimination performance on the basis of labeling data (Wood, 1976). The convergence between the notion of auditory discontinuities and categorical perception has been attractive to many investigators. Brown and Sinnott (chap. 12, this volume) place significant emphasis on categorical perception, which they refer to as the phoneme boundary effect, and they review nonhuman animal data for several speech contrasts and several species in search of putative sensory discontinuities.

Kuhl and Miller (1975, 1978; see also Kuhl, 1981) demonstrated that chinchillas (*Chinchilla laniger*) “label” and discriminate voiced ([b], [d], [g]) from voiceless ([p], [t], [k]) stop consonants in a fashion remarkably like that found for human listeners with the same stimuli—fulfilling all three of the preceding criteria. Consistent with claims made by Pisoni (1977) and by Jusczyk and his colleagues (Jusczyk, Pisoni, Walley, & Murray, 1980; Jusczyk, Rosner, Reed, & Kennedy, 1989), Kuhl and Miller (1975, 1978) argued that convergence of human and nonhuman data may be the result of real homology—presumably a common limit on the ability to temporally resolve the onset of aspiration energy and onset of periodic energy.

Sinex and McDonald (1988, 1989) later investigated neural responses to voiced and voiceless stop consonants in the auditory nerve (AN) of the chinchilla. When Sinex, McDonald, and Mott (1991) measured magnitude and variability of AN discharges in response to these stop consonants, they found that responses were larger, and variability was smaller, for stimuli near

the labeling boundaries for humans and chinchilla. They hypothesized that enhanced discrimination in perception tasks may be related to this increased neural resolution. Single-unit responses also have been measured in chinchilla inferior colliculus (Chen, Nuding, Narayan, & Sinex, 1996) as well as in auditory cortex in cat (Eggermont, 1995) and monkey (Steinschneider, Schroeder, Arezzo, & Vaughan, 1995). At the level of the cortex, a robust response to onset of periodic energy in voiceless stops was generally not detected until the time between syllable onset and onset of periodic energy reached a duration comparable to that typically found for identification boundaries in human listeners.

Early demonstrations by Kuhl and Miller were deeply troubling for proponents of human-specific processes for perceiving speech. More generally, however, these findings launched the still-pervasive suggestion that many or most contrasts between speech sounds exploit auditory discontinuities. By such a view, languages use distinctions for which modest changes in acoustic output result in disproportionately large perceptual consequences (Kuhl, 1986, 1988, 1993; Kuhl & Miller, 1978; Kuhl & Padden, 1982, 1983; Stevens, 1989). In Kuhl's (1993) formulation of the native language magnet theory, she suggested that gross phonetic categories are separated by "natural boundaries" produced by general auditory processes common among related species. The best support for this claim comes from the chinchilla studies cited earlier.

Here and there, modest evidence for other auditory discontinuities has been collected. Kuhl and Padden (1983) found that monkeys were better at discriminating two-formant consonant-vowel syllables (CVs) when members of stimulus pairs were more likely to be labeled differently ([b] vs. [d], or [d] vs. [g]) by adult human listeners. Dooling, Best, and Brown (1995) found that budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*) showed enhanced discrimination performance for pairs of CVs that are labeled differentially as [la] and [ra] by adult human listeners.

Brown and Sinnott (chap. 12, this volume) describe the generally conflicting findings for distinctions between liquids [r] and [l], and between stops and semivowels ([b] and [w]). They also describe the difficulty of interpreting findings by Kuhl and Padden (1983) for impoverished two-formant stimuli nominally varying from [bæ] to [dæ] to [gæ]. Overall, the search for homologous auditory predispositions for distinctions between speech sounds has not yielded overwhelming evidence that any speech-relevant sensory discontinuities exist for consonants beyond the evidence for voicing, as described earlier.

In contrast to Brown and Sinnott (chap. 12, this volume), we are skeptical about prospects for auditory discontinuities underlying speech distinctions in general. Searches for discontinuities with human and nonhuman subjects are unlikely to prove fruitful. There are several reasons why this is so.

First, categorical perception tasks with human listeners hardly serve to illuminate sensory discontinuities. Using signal detection analyses, Macmillan, Kaplan, and Creelman (1977) showed that differences in "sensitivity" depend more on particular methodology than on stimulus differences. Traditional use of two response alternatives (e.g., [ba], [pa]) makes for an insensitive identification measure, and habitual use of 2IAX, ABX, or AXB discrimination tasks consistently underestimates real sensitivity and enhances influences of memory. Because so much of speech perception depends on experience with distributions of speech sounds in one's language, it is not surprising that the addition of memory variables helps to predict performance on categorical perception tasks (Macmillan, 1987; Pisoni, 1973). When associative memory is simulated using neural network models (e.g., Anderson, Silverstein, Ritz, & Jones, 1977; Damper & Harnad, 2000), it appears that signature response patterns for categorical perception may be little more than an emergent property of any general learning system following exposure to realistic distributions of speech sounds.

To the extent that categorical perception is a consequence of experience with a structured environment with reliable stimulus patterns, one would expect categorical perception to exist for stimuli other than speech and for modalities other than audition. Categorical perception has been reported for musical intervals (Burns & Ward, 1974, 1978; Smith, Kemler Nelson, Grohskopf, & Appleton, 1994) and tempered triads (Locke & Kellar, 1973). Visually, humans categorically perceive human faces (Beale & Keil, 1995) and facial expressions (Calder,

Young, Perrett, Etcoff, & Rowland, 1996; De Gelder, Teunisse, & Benson, 1997; Etcoff & Magee, 1992), as well as cow faces morphed gradually to monkey faces (Campbell, Pascalis, Coleman, & Wallace, 1997). When human observers are trained with artificial categories, they gain acquired distinctiveness—increased perceptual sensitivity for items that are categorized differently (Goldstone, 1994). When monkeys are trained to respond differentially to clear examples of cats versus dogs (novel categories for monkeys), behavioral responses to stimuli along a morphed cat–dog series exhibit sharp crossovers at the series midpoint, and neural responses in prefrontal cortex correspond to these behavioral changes (Freedman, Riesenhuber, Poggio, & Miller, 2001).

The fact that categorical perception appears to be more related to perceptual experience than to sensory discontinuities cannot, by itself, imply nonexistence of auditory discontinuities. However, with the exception of voice-onset time (VOT), no strong case has ever been made for a sensory discontinuity underpinning a contrast between consonant or vowel sounds. In addition, the multitude of differences between speech sounds used across languages serves as evidence that it is unlikely that most contrasts are founded on sensory discontinuities. Even at the surface level, the implication is either that there would have to be a great many sensory discontinuities to accommodate so many distinctions, or that most distinctions are not predicated on sensory discontinuities.

OTHER AUDITORY INSIGHTS

Even though sensory discontinuities may not provide much of the explanation for speech perception, nonhuman animal models continue to play an essential and irreplaceable role in understanding raw auditory capacity as it relates to speech perception.

Studies by Lotto and his colleagues (Lotto & Kluender, 1998; Lotto, Kluender, & Holt, 1997) illustrate how data from nonhuman animals can illuminate auditory interactions that are central to perception of fluent speech. Their work addresses a ubiquitous challenge to speech researchers—perceptual accommodation of coarticulation. *Coarticulation* refers to the spatial and temporal overlap of adjacent articulatory activities. This overlap in production is reflected in the acoustic signal by severe context dependence; acoustic information specifying one speech sound varies substantially depending on surrounding consonants and vowels.

For example, one acoustic feature that contributes to perception of [d], particularly as contrasted with perception of [g], is the onset frequency and frequency trajectory of F_3 . In the context of following [a], a higher F_3 onset encourages perception of [da], whereas a lower onset results in perception of [ga]. Preceding context has considerable effect on the acoustic realization of the following consonant segment in fluent connected speech. The onset frequency of the F_3 transition varies as a function of the preceding consonant in connected speech. For example, F_3 -onset frequency for [da] is higher following [a] in [alda] relative to when following [ar] in [arda]. This pattern is due to the assimilative nature of coarticulation; the offset frequency of F_3 is higher for [a] (owing to a more forward place of articulation) and lower for [ar]. Schematic spectrograms depicting kinematic consequences of articulatory dynamics are depicted in Fig. 13.1.

Perception of [da] and [ga] has been shown to be affected by the composition of preceding acoustic information in a fashion that accommodates these patterns in production. For a series of synthesized CVs varying in onset characteristics of the third formant (F_3) and varying perceptually from [da] to [ga], individuals are more likely to perceive [da] when preceded by the syllable [ar], and to perceive [ga] when preceded by [a] (Mann, 1980). The effect has been found for speakers of Japanese who cannot distinguish between [l] and [r] (Mann, 1986) and for 4- to 5-month-old infants (Fowler, Best, & McRoberts, 1990). The important point is that, for the very same stimulus with F_3 onset intermediate between [da] and [ga], the percept is altered as a function of preceding context. Listeners perceive speech in a manner that seems to imply sensitivity to the compromise between production of neighboring sounds.

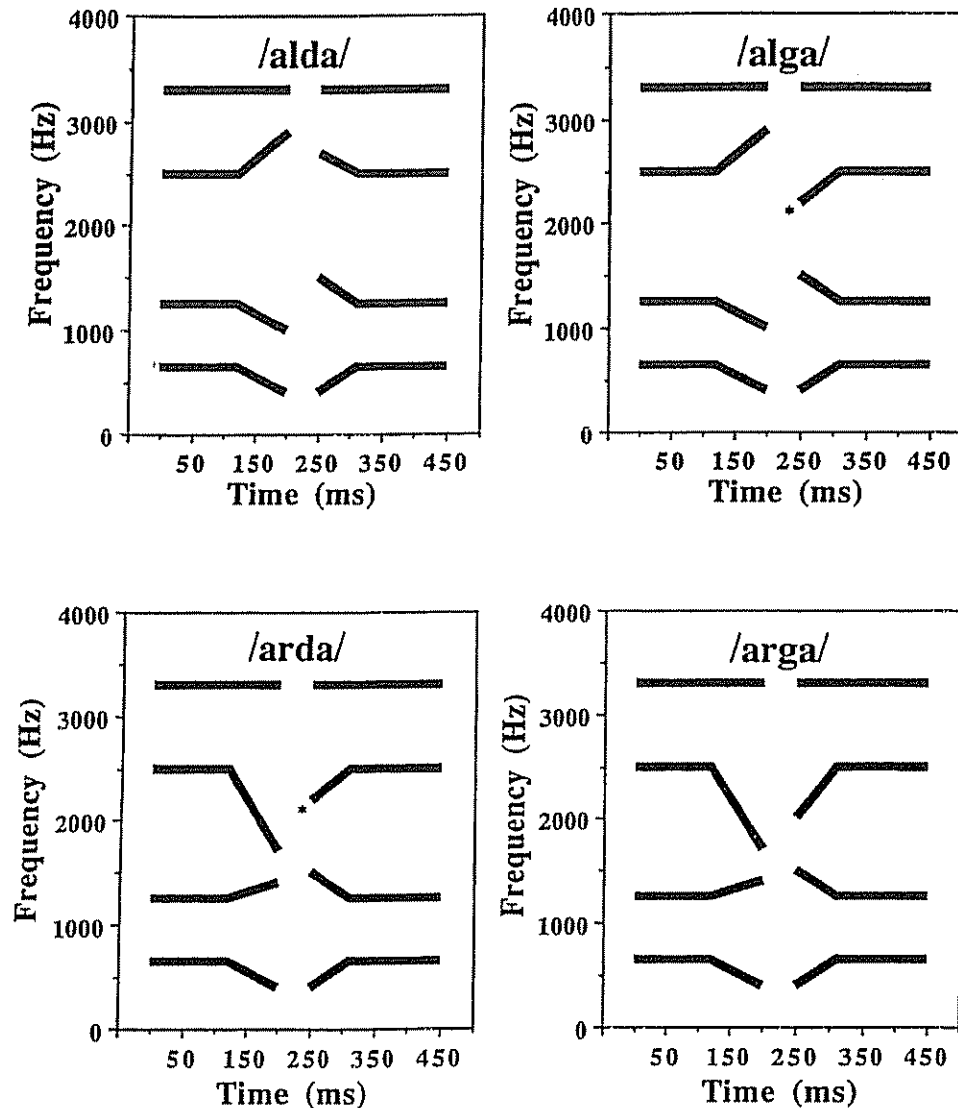


FIG. 13.1. Schematic spectrograms of the disyllables [alda], [alga], [arda], and [arga] portraying acoustic assimilation as a consequence of coarticulation. For the full series used by Lotto, Kluender, and Holt (1997), the change from [da] to [ga] was signaled by F_2 onset frequency. Note that the CV perceived as "ga" in [alga] (upper right) is acoustically identical to the CV perceived as "da" in [arda] (lower left).

Lotto et al. (1997) conducted an experiment to discover whether animals would exhibit a similar effect of preceding context. Japanese quail were trained to peck a key when presented exemplars of either the syllable [da] or the syllable [ga] and to refrain from pecking to [ga] or [da], respectively. Birds then were presented with novel ambiguous CVs preceded by either [al] or [ar]. The task was essentially the same as that used by Mann (1980) with human listeners. The birds "labeled" CVs as [da] or [ga]. All four quail displayed a significant shift in peck rates across the change in preceding liquid. As seen in Fig. 13.2, two [da]-positive birds pecked substantially more to CVs preceded by [ar] while [ga]-positive birds pecked more to CVs preceded by [al]. This is the same shift Mann (1980) showed for human listeners.

It is unlikely that quail explicitly or tacitly compensated for acoustic effects of coarticulation arising from human vocal tracts, as they had no experience with patterns of coarticulation. Lotto and Kluender (1998) investigated how the parallel patterns of results for humans and quail could arise from general (and homologous) auditory processes. They noted that the contextual effects of preceding [ɾ] or [l] could be described as examples of spectral contrast. If one considers F_3 frequencies for the CVs and the [al] and [ar] stimuli, one sees that performance of animals (and humans) can be described as contrastive. Quail trained to peck to CVs with low F_3 -onset frequencies ([ga]-positive) pecked more to intermediate values of F_3 -onset frequency (novel ambiguous stimuli) when CVs were preceded by a syllable with a high F_3 offset ([al]). Quail trained to peck to CVs with high F_3 -onset frequencies ([da]-positive) pecked more to intermediate values of F_3 -onset frequency when CVs were preceded by a syllable with a low-frequency F_3 -offset ([ar]).

Avian data suggest that this contrastive pattern may be a ubiquitous product of auditory systems and is not a pattern of behavior that has evolved in humans to deal specifically with coarticulation. If this simple spectral contrast hypothesis is true, then analogous effects may obtain when energy other than speech precedes the [da-ga] syllables. Lotto and Kluender (1998) found that human identification functions shifted when CVs were preceded by a frequency-modulated (FM) sine wave glide mimicking the F_3 transition of [al] or [ar], and even by a constant-frequency sine wave set at the frequency of F_3 offset for [al] or [ar].

This pattern of results for [da] and [ga] in VCCVs consequently has been demonstrated for medial vowels in CVCs and medial stops in VCVs. Listeners' perception of vowels varying perceptually from [ɛ] to [I] in CVC syllables are more likely to be heard as [ɛ] (lower F_2) following [d] (higher F_2) and as [I] following [b] (Holt, Lotto, & Kluender, 2000). Analogously, listeners are more likely to

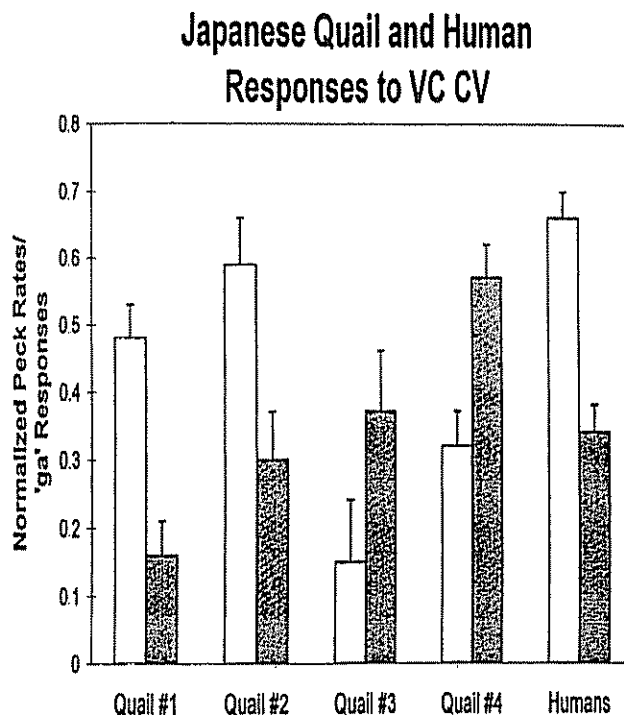


FIG. 13.2. The first eight bars of the histograms indicate response strength (peck rates) for Japanese quail hearing CV syllables intermediate between [da] and [ga] endpoints following [al] (unfilled bars) and [ar]. Quail 1 and 2 were reinforced for pecking to [ga]. Quail 3 and 4 were reinforced for pecking to [da]. Data are from Lotto et al. (1997). Human data are [ga] responses to the same stimuli from Lotto and Kluender (1998) for a study patterned closely after Mann (1980).

perceive stimuli from a series varying from [b] (lower F_2) to [d] (higher F_2) as [b] following [i] (high F_2) and as [d] following [u] (low F_2 ; Coady & Kluender, 2002; Holt, 1999; Holt & Lotto, 2002). In general, these contrast effects are maintained following both speech and nonspeech energy to the extent that preceding acoustic energy provides contrastive spectral composition.

Interpretation of these contrastive effects and their relation to perception of coarticulated speech can be offered at several levels. At a more abstract (and less explanatory) level, these effects can be seen as emblematic of a system that adheres to quite general constraints on sound change over time (Kluender, 1991; Lotto et al., 1997). Due to mass and inertia, physical systems tend to be assimilative. The configuration of a system at time, t , is significantly constrained by its configuration at time $t-1$. Vocal tracts are bound by these same physical constraints; kinematic acoustic consequences of coarticulation are assimilative. Because very rapid change is the exception for physical systems, signs of change are emphasized through processes of perceptual contrast. The resulting symmetry of speech production and perception is not serendipitous. It is a consequence of organisms having evolved to interact most generally with physical systems that are constrained across time.

Sensitivity to change, of course, is not necessarily restricted to complementarity between physics of sound production and constraints on perception of sound. It is both true and fortunate that sensorineural systems respond to change and little else (Kluender, Coady, & Kiefte, 2003). Perceptual systems do not record absolute level be it amplitude and frequency in acoustics, or luminance and wavelength in optics. This fact has been demonstrated in every sensory domain. Physiologically, sensory encoding is always relative. This sacrifice of absolute encoding has enormous benefits along the way to maximizing information transmission. Biological sensors have impressive dynamic range given their evolution via borrowed parts (e.g., gill arches becoming middle ear bones). However, biological dynamic range always is a small fraction of the physical range of absolute levels available in the environment, as well as in the perceptual range essential to organisms' survival. This is true whether one is considering optical luminance or acoustic pressure. The beauty of sensory systems is that, by responding to relative change, a limited dynamic range adjusts to maximize the amount of change that can be detected in the environment.

The simplest way that sensory systems adjust dynamic range to maximize sensitivity to change is via adaptation. Following minimal stimulation, a sensory stimulus triggers a strong sensation. However, when sustained sensory input does not change over time, constant stimulation loses impact. This sort of sensory attenuation due to adaptation is ubiquitous, and has been documented in vision (Riggs, Ratliff, Cornsweet, & Cornsweet, 1953), audition (Hood, 1950), taste (Urbantschitsch, 1876, cited in Abrahams, Krakauer, & Dallenbach, 1937), touch (Hoagland, 1933), and smell (Zwaardemaker, 1895, cited in Engen, 1982). There are more mechanisms supporting this sensitivity to stimulus change, but perception of any object or event is always relative—critically dependent on its context.

Candidate auditory mechanisms—which do the real work of providing spectral contrast—may include well-established processes of sensory adaptation. Delgutte (1996) noted that peaks in AN discharge rate correspond to spectro-temporal regions that are rich in information, and that adaptation increases the temporal resolution with which onsets are represented. Most important to issues related to coarticulation, adaptation enhances spectral contrast between successive speech segments. This enhancement arises because a fiber adapted by stimulus components close to its characteristic frequency (CF) is relatively less responsive to subsequent energy at that frequency, whereas stimulus components not present immediately prior are encoded by fibers that are unadapted.

Simple adaptation, at least at the lowest levels of the auditory system, can only provide a partial explanation for the spectral contrast effects found for speech and hybrid nonspeech–speech stimuli. In part, this is because the time course for contrastive effects for speech extend over longer durations than typically found for adaptation of AN fibers (Holt & Lotto, 2002). Further, effects for both speech (Holt & Lotto, 2002) and nonspeech (Lotto, Sullivan, & Holt, 2003) are maintained even when the first contrastive sound is presented to the opposite ear. These findings suggest that adaptation or adaptation-like effects important to perception of fluent speech must also involve loci beyond

the AN. Taken together, these findings with human and nonhuman subjects suggest that one of the most pervasive problems for models of speech perception may be at least partially explained via general auditory processes. Cross-species data using both speech and nonspeech stimuli provide converging sources of evidence that appear compelling.

Before leaving this section, the importance of animal models bears emphasis. Data from nonhuman subjects—in both behavioral and neurophysiological studies—played a special role in suggesting the appropriate final analysis. Although it may have been tempting to draw strong conclusions only from data from human participants hearing nonspeech mimics of speech signals (e.g., Lotto & Kluender, 1998), such demonstrations have not proven universally convincing. Kuhl (1978, 1986) suggested that nonspeech stimuli may be processed by “speech-specific” mechanisms with rather broad application. Pisoni (1987) argued that “it may be that nonspeech stimuli that are sufficiently speech-like are processed by central processes involved in speech perception, even when the listener is not aware of the speech-likeness of the stimuli” (p. 266). For example, if the spectral temporal structure of a nonspeech stimulus (e.g., FM glide) corresponds sufficiently with the structure of speech (e.g., F_3 transition), “speech-specific” processes may not be so finely tuned as to register the nonspeech–speech difference. Suga (chap. 11, this volume) suggests that neurons putatively specialized to respond to biologically significant sounds (i.e., feature detectors) should more appropriately be considered specialized information-bearing parameter filters (IBFs) because they appear somewhat broadly tuned. This appraisal would be consistent with Kuhl’s and Pisoni’s reluctance to find data from studies using human participants and nonspeech stimuli wholly compelling.

Another possible (and related) objection to drawing strong conclusions in the absence of nonhuman animal models is that the extremely overlearned nature of speech can affect perception of nonspeech sounds that mimic particular aspects of the speech signal. When one considers, for example, connectionist simulations of learning with a bank of filters serving as input, output is indifferent to the precise nature of the energy within a frequency band. In a simulation following training with speech sounds, nonspeech signals sharing similar acoustic properties would be processed as if they were speech signals. It is not unreasonable to suggest that, for human listeners, abundant experience with speech could have similar consequences. The nonhuman animal model on the other hand, can be deprived of experience with speech signals and, thus, serves as the only certain measure of general auditory contributions.

THE INITIAL STATE

Much of the emphasis thus far has been on the use of nonhuman animals to eliminate potential confounding influences of experience with speech, and the case has been made that nonhuman subjects hold clear advantages over adult human participants in this respect, even for studies employing nonspeech stimuli. Historically, however, most studies designed to titrate effects of audition from effects of experience have been conducted with human infants. In this context, it is a curious fact that many more studies of human infant auditory perception have been carried out using speech stimuli than with all other types of stimuli combined (e.g., tones, noises, nonspeech environmental sounds).

Nearly three decades of studies document the impressive abilities of human infants, some less than 1 week old, to discriminate a wide variety of consonants and vowels (see Eimas, Miller, & Jusczyk, 1987; Kluender, 1994; Kuhl, 1987, for comprehensive reviews). A plethora of positive findings indicate that infants have the discriminative capacity necessary for most or all of the speech distinctions they will need to use in their language. Early infant auditory abilities appear to be quite well developed, and by 3 months of age, the human auditory system is nearly adult-like in absolute sensitivity and frequency-resolving power within the frequency range of most speech sounds (Werner & Bargones, 1996).

Although the majority of speech studies were conducted with infants 4 to 6 months of age, some have been conducted with newborns. The general impetus for most of these studies has been to evaluate the degree to which specialized processes may prepare the human infant for language acquisition

in a fashion unavailable to nonhuman organisms. What appears to be the case, however, is that barring imposition of impractical and unethical limits on early experience, there is no way to test the human infant's perception of speech sounds unfettered by the effects of language experience. By the time French infants are 4 days old, not only can they discriminate French from other languages (e.g., Russian), but they also exhibit a preference for French as spoken in the birthing hospital and in their parent's home (Mehler et al., 1988).

One might attribute this ability to experience gained during those first few postpartum days. If this were so, the sole challenge to researchers would be to use methods that precede extrauterine experience and assess only a pristine perceptual system. The problem, however, is that experience begins much earlier. Mehler et al. (1988) obtained the same general results when French and Russian passages were low-pass filtered (400 Hz). Although this manipulation restricted the signal mostly to those aspects that convey prosody (i.e., fundamental frequency [f_0], relative durations, and amplitude), filtering also made the signal more like that available prenatally.

There appears to be significant prenatal exposure to environmental sounds, including speech, through the abdominal walls (e.g., Griffiths, Brown, Gerhardt, Abrams, & Morris, 1994; Lecanuet, Granier-Deferre, Cohen, Le Houezec, & Busnel, 1986). In fact, measurements using cardiac deceleration as an indicator of discrimination in a habituation task indicate that late-term fetuses can discriminate vowel sounds (Lecanuet et al., 1986). This prenatal experience with speech sounds appears to have considerable influence on subsequent perception, as newborns prefer their mother's voice (DeCasper & Fifer, 1980). Most telling is the finding that newborns prefer hearing particular speech passages (children's stories) that were read aloud by their mothers during the third trimester (DeCasper & Spence, 1986).

The unfortunate (or at least inconvenient) conclusion from this is that, even for the newborn, traditional studies of infant speech perception are too compromised by effects of prenatal experience to afford unconfounded evaluation of preparedness for speech perception. It long has been known that adult perception of speech contrasts is extremely dependent on experience with a particular language (e.g., [r] vs. [l] by native Japanese listeners; Miyawaki et al., 1975). Worse, it appears that there is little hope that performance by human participants of any age can reveal raw auditory capacities for processing speech.

Studies with nonhuman animals hearing speech and other complex sounds offer the investigator the only way to behaviorally evaluate default sensory capacities as they pertain to speech. Over more than two decades, there have been a good many demonstrations that animals can distinguish individual human speech sounds with facility. For example, differences between vowel sounds present no apparent difficulty to animals (Burdick & Miller, 1975; Kluender & Diehl, 1987; Kluender, Lotto, Holt, & Bloedel, 1998). Budgerigars (*Melopsittacus undulatus*; Dooling, Okanoya & Brown, 1989), chinchilla (*Chinchilla laniger*; Kuhl, 1981; Kuhl & Miller, 1975, 1978), macaques (*Macaca mulata*; Kuhl & Padden, 1982), and Japanese quail (*Coturnix coturnix japonica*; Kluender, 1991; Kluender & Lotto, 1994) all are adept at distinguishing voiced ([b], [d], [g]) from voiceless ([p], [t], [k]) stop consonants. Macaques (Kuhl & Padden, 1983) and Japanese quail (Kluender & Diehl, 1987; Kluender et al., 1987; Lotto et al., 1997) can distinguish place of articulation (labial [b], alveolar [d], velar [g]) for stop consonants. Budgerigars and zebra finches (Dooling et al., 1995) and macaques (Sinnott & Brown, 1997) can distinguish the liquids [l] and [r], an ability sharply attenuated for Japanese adults.

More recent work has established striking similarities between human and nonhuman responses to extended prosodic qualities of speech. Ramus and colleagues (Ramus, Hauser, Miller, Morris, & Mehler, 2000), for example, compared discrimination of Dutch and Japanese sentences by human newborns and cotton-top tamarins. Performance was remarkably similar for newborns and the New World monkeys, including the finding that both groups were better at discriminating sentences when presented forward versus backward. Given the opportunity to avoid confounding effects of language experience, animal models—as homologues—offer the only unadulterated measure of auditory capacities underlying perception of human speech.

EXPERIENCE AND SPEECH PERCEPTION

To begin, it must be understood that the ability to discriminate speech sounds is not synonymous with the functional use of speech sounds. Whereas a relatively complete understanding of the auditory representation of speech sounds is an essential step in understanding speech perception by itself, mapping acoustic energy onto either neural representations or some more abstract perceptual space falls short. There are many variations on the speech signal, some linguistically meaningful, and some not. In this multidimensional space, experience plays a critical role in parceling speech sounds into functional classes with linguistic significance (i.e., phonetic categories). The fact that different languages partition the domain of possible speech sounds differently implies that functional use of speech sounds depends on linguistic experience. Different languages use different subsets from the broad assortment of sounds used linguistically by humans. Acoustic differences that are functionally neglected for one language can be communicatively critical for listeners of another language.

A related concern arises because the speech signal is so rich and complex. Contrasts between speech sounds are signaled by a broad array of physical acoustic differences, none of which provides a necessary and sufficient cue to identity. As such, it is generally accepted that consonants and vowels do not have invariant identifying properties. More positively, such multiplicity lies at the heart of redundancy and robustness of speech as a communication channel. Although there have been a good many efforts to identify invariant cues (e.g., Blumstein & Stevens, 1979; Stevens & Blumstein, 1981), these efforts have not been broadly successful. By accepting the requirement of multiple acoustic or auditory attributes, it becomes clear that even the most definitive description of auditory representations and processes will fall short of explaining how multiple attributes conspire toward distinguishing speech sounds. Fortunately, what biological systems do well is use multiple sources of inconsistent or noisy data toward some perceptual end. Most contemporary models of learning and neural organization are designed to capture just this fact.

Here, it is suggested that integration of multiple stimulus attributes is the result of effects of experience with speech sounds. A full model of human speech perception will require incorporation of processes through which experience with speech gives rise to functional linguistic distinctions. These are not troubling developments. The hallmark of primate behavioral development may be plasticity—well-supported by a generous cortical endowment. At the extreme end of this developmental continuum one finds humans. Humans are the most plastic of species, and the temporal lobe (along with frontal lobe) is considered to be among the most plastic structures of the mammalian brain.

Echoed again is the rejection of simple signaling systems and biosonar as apt analogies—let alone homologies—for human speech communication. In human communication, there is no compelling evidence for broad stereotypy emblematic of hard-wired systems. Plastic perceptual processes molded by experience better match the extant facts of speech perception within and across languages. Suga (chap. 11, this volume) recommends dorsal and possibly medial subdivisions of the medial geniculate body (MGB) as loci for separate and specialized processing of speech sounds by humans. It is difficult to argue relative merits of respective brain loci given the present dearth of knowledge—particularly with respect to MGB (Budinger & Heil, chap. 7, this volume; Clarey, Barone, & Imig, 1992; Winer, 1992). However, it can be stated with confidence that dorsal and medial MGB must be exquisitely sensitive to sensory experience if the auditory thalamus is to play a major part, beyond being merely a pathway, in the perception of speech tuned to the sounds of particular languages.

Whereas nonhuman animals do not have the cognitive potential of humans, more positively, nonhuman subjects provide homologues for simpler types of perceptual change via experience. Embracing nonhuman models as homologues carries the explicit assumption that underlying processes supporting perceptual development are relatively consistent across species. In addition to the virtues of parsimony, this conservative claim includes an appreciation that the very virtue of plastic processes is adaptiveness across varying environments. This is not to say that simpler organisms share the full human potential for change as a function of experience. As will be seen, the more austere neural potential of nonhumans can be a virtue when developing parsimonious models. In studies of the

effects of experience for speech perception, nonhuman subjects are used in a different way than previously described. Instead of using nonhuman animal subjects to eliminate confounding influences of experience, nonhuman subjects can be used in ways that permit studying perceptual learning while maintaining exquisite control over experience.

For a while now, there have been suggestions that one may be able to make do with quite simple processes in accounting for perceptual learning of functional equivalence classes of speech sounds. Kluender et al. (1987) found that Japanese quail could master the ostensibly complex mapping between multiple acoustic attributes and a response consistent with human perception of alveolar stop consonants. Despite the fact that quail have brains about the size of an almond, performance generalized with facility to previously unheard syllables beginning with [d].

Sussman and his colleagues (e.g., Sussman, 1989, 1994, 2002; Sussman, Fruchter, & Cable, 1995) have shown that place of articulation for utterance-initial stops, as characterized by covariation between onset frequency of F_2 and quasi-steady-state frequency of F_2 in the vowel portion of CVs, can be captured reasonably well by simple linear operations. Sussman and colleagues (Sussman, 2002; Sussman, Fruchter, Hilbert, & Sirosh; 1998) argue—by analogy to bat and owl neural encoding—that if human listeners are sensitive to this covariation, it could be taken as evidence that analogous adaptations for speech sounds may be present in humans. For us, this linear mapping is not taken as an argument for specialized processes. Instead, acoustic products of articulation, covariance between F_2 onset and F_2 vowel, may be ideal grist for the simplest sorts of perceptual learning (Kluender, 1998). The preceding quail data stand in support of learnability over specialization, as quail are unlikely genetic recipients of specialized processes for perception of speech.

Consider a recent study by Holt, Lotto, and Kluender (2001). They investigated the well-established fact that, for many languages including English, fundamental frequency (f_0) and voicing tend to covary. Talkers produce voiceless stops such as [p] with higher f_0 (at onset of periodic energy) than voiced stops. English listeners perceive voicing in a fashion consistent with these articulatory facts. When listeners identify stimuli that vary perceptually from voiced to voiceless (e.g., from [b] to [p]), they report hearing more sounds as voiced [b] when f_0 is lower (e.g., Castleman & Diehl, 1996; Chistovich, 1969; Haggard, Ambler, & Callow, 1970; Haggard, Summerfield, & Roberts, 1981; Kohler, 1984; Whalen, Abramson, Lisker, & Mody, 1993). Some investigators (Diehl & Kluender, 1989; Kingston & Diehl, 1994) have suggested that talkers covary f_0 and voicing because these two acoustic qualities conspire auditorily in as much as voiced consonants have greater low-frequency energy and a lower f_0 enhances this distinctive auditory quality.

Holt and her colleagues tested the alternative hypothesis that the perceptual effect of listeners hearing more voiced stops when f_0 is lower could instead be due to listeners' experience with the reliable covariance between f_0 and voicing when listening to speech. Because human listeners share significant experience with this covariation, Holt et al. (2001) used Japanese quail as subjects to precisely control effects of perceptual learning. They trained three groups of quail to respond differentially to voiced versus voiceless stops. Quail were trained with either of three patterns of f_0 voicing. Voicing and f_0 were varied in the natural pattern (voiced, low f_0), in an inverse pattern (voiced, high f_0), or in a random pattern (no systematic f_0 covariance). Birds trained with stimuli from the latter condition (no voicing/ f_0 covariance) exhibited no effect of f_0 on responses to novel stimuli intermediate between voiced and voiceless endpoints. For the other two groups, with systematic variation between voicing and f_0 , the birds' patterns of responses followed the experienced pattern of covariance. Complete control over animals' experience with speech sounds permitted the conclusion that f_0 may not exert an obligatory influence on perception of voicing and that, instead, the patterns of performance found with human listeners are more likely due to the learnability of covariance among acoustic characteristics of speech.

Relatively simple general processes of perceptual learning also may account for what has been termed the perceptual magnet effect (PME; Kuhl, 1991, 1993, 1994). Most broadly, the PME refers to several observations beginning with the finding that some acoustic instances of vowel sounds are perceptually more compelling than others. This has been related to the analogous finding in hu-

man concept formation whereby categories can be construed as being structured about a prototype with exemplars more like the prototype being judged as “better” examples of the category. Suga (chap. 11, this volume) carefully explicates how PME claims may be given neural interpretation within his IBE framework.

Kluender et al. (1998) used European starlings (*Sturnus vulgaris*) to evaluate the extent to which graded perception of vowels found for human listeners can be explained by general processes of perceptual learning that are sensitive to statistical distributions of experienced sounds. They trained eight starlings to discriminate vowel tokens drawn from stylized distributions either of English vowel sounds [i] and [I], or of Swedish vowel sounds [y] and [u], as depicted in Fig. 13.3. Following training, responses to novel stimuli drawn from these distributions indicated that the starlings generalized with facility to novel examples. Responses were well accounted for on the bases of F_1 and F_2 values, as well as by distance from the centroid of these distributions of vowel sounds, thus manifesting graded structures like those often taken to imply the existence of category prototypes. Moreover, starling response rates corresponded closely to adult human judgments of “goodness” for English vowels [i] and [I] ($r = .999$ across full two-vowel task, mean $r = .714$ within individual distributions). Quite telling was the fact that a simple linear association network model trained with vowels drawn from the birds’ training set captured 95% of the variance in the birds’ response rates for novel vowel tokens. Human judgments of goodness seem to be quite well accommodated by general perceptual learning processes given exposure to probability-density distributions like those heard by human listeners.

Swedish/English Vowel Distributions

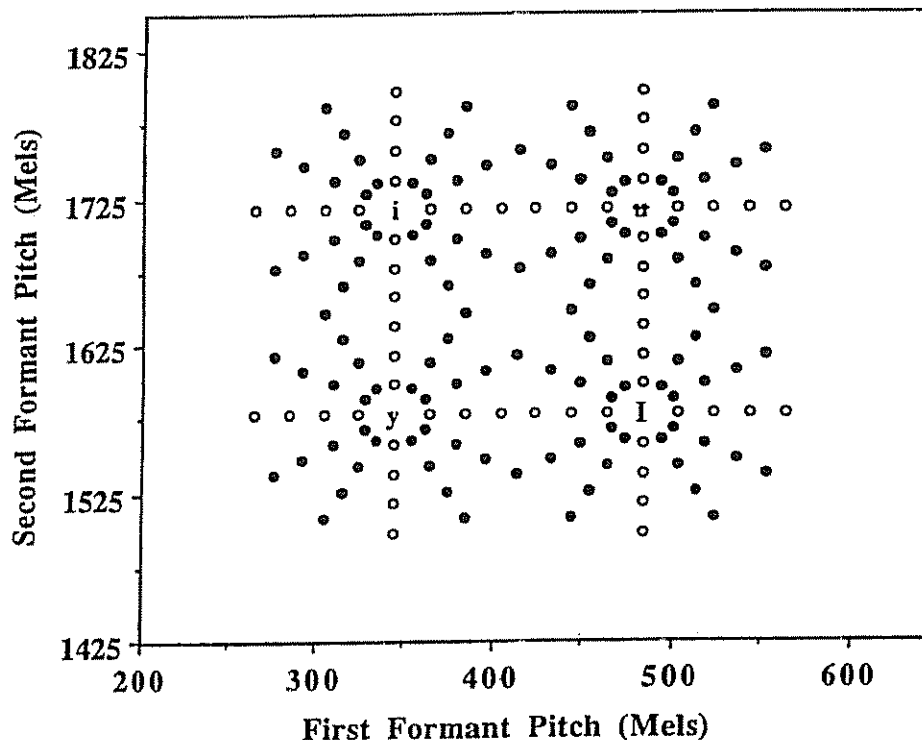


FIG. 13.3. Mel-scaled plot of 196 synthetic vowel stimuli representing equal 49-token distributions of the English vowels [i] and [I] and the Swedish vowels [y] and [u]. Filled symbols represent stimuli used in training. Unfilled circles and the symbols [i], [I], [y], and [u] (centroids) correspond to stimuli withheld until testing and presented without reinforcement.

It appears that there is no reason to believe that general principles of learning, whether instantiated in nonhuman animals, in computational models using covariance matrices or connectionist networks, or in real neural networks, will be adequate to explain the structure of linguistic functional equivalence classes for speech sounds.

One type of information that will be critically important in the development of adequate models for perceptual experience with speech will be richer characterization of the input to the process. Part of developing a better understanding of the input will take the form of more accurate sensory and neural descriptions of speech sounds as they pass through the auditory system (e.g., Jenison, 1991; Jenison, Greenberg, Kluender & Rhode, 1991). Efforts using nonhuman animal models to describe auditory transformations will prove very helpful in the future, and these transforms could provide valuable "front ends" to computational simulations of plastic perceptual organization.

In the meantime, nonhuman animal models of perceptual experience (unlike computers) have the virtue of coming to the laboratory with their auditory transformations in place. In this role, nonhuman animal models play the role of homology with respect to both foundational auditory sensory processes and more plastic processes of perceptual development. One facile way to divide the problem of explaining human speech perception in general is to provide a description of the sensory representation of speech in the auditory system and also to provide a model of how experience with speech sounds shapes perceptual performance with speech. In each of these roles, the nonhuman animal provides an invaluable vehicle to do so.

ACKNOWLEDGMENTS

Preparation of this chapter supported by NIDCD DC04072. Written in 2002.

REFERENCES

- Abbs, J. H., & Sussman, H. M. (1971). Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. *Journal of Speech and Hearing Research, 14*, 23–36.
- Abrahams, H., Krakauer, D., & Dallenbach, K. M. (1937). Gustatory adaptation to salt. *American Journal of Psychology, 49*, 462–469.
- Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84*, 413–451.
- Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57*, 217–239.
- Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America, 66*, 1001–1017.
- Burdick, C. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America, 58*, 415–427.
- Burns, E. M., & Ward, W. D. (1974). Categorical perception of musical intervals. *Journal of the Acoustical Society of America, 55*, 456.
- Burns, E. M., & Ward, W. D. (1978). Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63*, 456–468.
- Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3*, 81–117.
- Campbell, R., Pascalis, O., Coleman, M., & Wallace, B. (1997). Are faces of different species perceived categorically by human observers? *Proceedings of the Royal Academy of London, 264*, 1429–1434.
- Castleman, W. A., & Diehl, R. L. (1996). Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics, 24*, 383–398.
- Chen, G.-D., Nuding, S. C., Narayan, S. S., & Sinex, D. G. (1996). Responses of single neurons in the chinchilla inferior colliculus to consonant–vowel syllables differing in voice onset time. *Auditory Neuroscience, 3*, 179–198.
- Chistovich, L. A. (1969). Variations of the fundamental voice pitch as a discriminatory cue for consonants. *Soviet Physics and Acoustics, 14*, 372–378.

- Clarey, J. C., Barone, P., & Imig, T. J. (1992). Physiology of thalamus and cortex. In A. N. Popper & R. R. Fay (Eds.), *The mammalian auditory pathway: Neurophysiology* (pp. 232–334). New York: Springer Verlag.
- Cleveland, J., & Snowdon, C. T. (1982). The complex vocal repertoire of the adult cotton-top tamarin (*Sanguinus oedipus*). *Zeitschrift für Tierpsychologie*, 58, 231–270.
- Coady, J. A., & Kluender, K. R. (2002). Importance of onset properties on spectral contrast for speech and other complex spectra. *Journal of the Acoustical Society of America*, 111, 2434.
- Damper, R. I., & Harnad, S. R. (2000). Neural network models of categorical perception. *Perception and Psychophysics*, 62, 843–867.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208, 1174–1176.
- DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, 9, 133–150.
- De Gelder, B., Teunisse, J.-P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion*, 11, 1–23.
- Delgutte, B. (1996). Auditory neural processing of speech. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 507–538). Oxford, UK: Blackwell.
- Diehl, R. L. (1981). Feature detectors for speech: A critical reappraisal. *Psychological Bulletin*, 89, 1–18.
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, 1, 121–144.
- Diehl, R. L., Kluender, K. R., & Parker, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception and Performance*, 11, 209–220.
- Dooling, R. J., Best, C. T., & Brown, S. D. (1995). Discrimination of full formant and sine wave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *Journal of the Acoustical Society of America*, 97, 1839–1846.
- Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced–voiceless distinction. *Perception and Psychophysics*, 46, 65–71.
- Eggermont, J. J. (1995). Representation of a voice onset time continuum in primary auditory cortex of the cat. *Journal of the Acoustical Society of America*, 98, 911–920.
- Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99–109.
- Eimas, P. D., Miller, J. L., & Jusczyk, P. W. (1987). On infant speech perception and the acquisition of language. In S. Harnad (Ed.), *Categorical perception* (pp. 161–195). Cambridge, UK: Cambridge University Press.
- Engen, T. (1982). *The perception of odors*. New York: Academic.
- Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, 44, 227–240.
- Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, 4, 258–267.
- Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception and Psychophysics*, 48, 559–570.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical perception of visual stimuli in the primate prefrontal cortex. *Science*, 291, 312–316.
- Frishkopf, L. S., & Goldstein, M. H. (1963). Responses to acoustic stimuli from single units in the eighth nerve of the bullfrog. *Journal of the Acoustical Society of America*, 35, 1219–1228.
- Gerhardt, H. C., & Rheinlander, J. (1982). Localization of an elevated sound source by the green tree frog. *Science*, 217, 663–664.
- Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123, 178–200.
- Griffiths, S. K., Brown, W. S., Gerhardt, K. J., Abrams, R. M., & Morris, R. J. (1994). The perception of speech sounds recorded within the uterus of a pregnant sheep. *Journal of the Acoustical Society of America*, 96, 2055–2064.
- Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, 47, 613–617.
- Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading *F* cues in the voiced–voiceless distinction. *Journal of Phonetics*, 9, 49–62.
- Hoagland, H. (1933). Quantitative aspects of cutaneous sensory adaptation I. *Journal of General Physiology*, 16, 911–923.
- Holt, L. L. (1999). *Auditory constraints on speech perception: An examination of spectral contrast*. Unpublished doctoral dissertation, University of Wisconsin, Madison, WI.

- Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the neural mechanisms of speech context effects. *Hearing Research, 167*, 156–169.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America, 108*, 710–722.
- Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America, 109*, 764–774.
- Hood, J. D. (1950). Studies in auditory fatigue and adaptation. *Acta Oto-Laryngology Supplement, 92*
- Hoy, R., Hahn, J., & Paul, R. C. (1977). Hybrid cricket auditory behavior: Evidence for genetic coupling in animal communication. *Science, 195*, 82–83.
- Jenison, R. L. (1991). *A dynamic model of the auditory periphery based on the responses of single auditory-nerve fibers*. Unpublished doctoral dissertation, University of Wisconsin, Madison, WI.
- Jenison, R. L., Greenberg, S., Kluender, K. R., & Rhode, W. S. (1991). A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory nerve fibers. *Journal of the Acoustical Society of America, 90*, 289–305.
- Jusczyk, P. W., Pisoni, D. B., Walley, A., & Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America, 67*, 262–270.
- Jusczyk, P. W., Rosner, B. S., Reed, M. A., & Kennedy, L. J. (1989). Could temporal order differences underlie 2-month-olds' discrimination of English voicing contrasts? *Journal of the Acoustical Society of America, 90*, 83–96.
- Katz, L. C., & Gurney, M. E. (1981). Auditory responses in the zebra finch's motor system for song. *Brain Research, 221*, 192–197.
- Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70*, 419–454.
- Kluender, K. R. (1991). Psychoacoustic complementarity and the dynamics of speech perception and production. *Perilus, 14*, 131–135.
- Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 173–217). San Diego, CA: Academic.
- Kluender, K. R. (1998). Locus equations reveal learnability. *Behavioral and Brain Sciences, 21*, 273.
- Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication, 41*, 59–69.
- Kluender, K. R., & Diehl, R. L. (1987, November). *Use of multiple speech dimensions in concept formation by Japanese quail*. Paper presented at the 114th meeting of the Acoustical Society of America, Miami, FL.
- Kluender, K. R., & Lotto, A. J. (1994). Effects of first formant onset frequency on [-voice] judgments: Results from general auditory processes not specific to humans. *Journal of the Acoustical Society of America, 95*, 1044–1052.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science, 237*, 1195–1197.
- Kluender, K. R., Lotto, A. J., Holt, L. L., & Bloedel, S. L. (1998). Role of experience for language-specific functional mappings of vowel sounds. *Journal of the Acoustical Society of America, 104*, 3568–3582.
- Kohler, K. J. (1984). Phonetic explanation in phonology: The feature fortis/lenis. *Phonetica, 41*, 150–174.
- Kuhl, P. K. (1978, May). *Perceptual constancy for speech-sound categories*. Paper presented at the N.I.C.H.D. Conference on Child Phonology: Perception, Production, and Deviation, Bethesda, MD.
- Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America, 70*, 340–349.
- Kuhl, P. K. (1986). Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology, 45*, 233–265.
- Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception* (Vol. 2, pp. 257–381). New York: Academic.
- Kuhl, P. K. (1988). Auditory perception and the evolution of speech. *Human Evolution, 3*, 19–43.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics, 50*, 93–107.
- Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. Schonon, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). The Hague, Netherlands: Kluwer.
- Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology, 4*, 812–822.

- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in the alveolar-plosive consonants. *Science*, *190*, 69-72.
- Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, *63*, 905-917.
- Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics*, *32*, 542-550.
- Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, *73*, 1003-1010.
- Lecanuet, J. P., Granier-Deferre, C., Cohen, C., Le Houezec, R., & Busnel, M. C. (1986). Fetal responses to acoustic stimulation depend on heart rate variability pattern stimulus intensity and repetition. *Early Human Development*, *13*, 269-283.
- Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineers*, *47*, 1940-1951.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1-36.
- Lieberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, *243*, 489-493.
- Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University Press.
- Locke, S., & Kellar, L. (1973). Categorical perception in a non-linguistic mode. *Cortex*, *9*, 353-369.
- Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception and Psychophysics*, *60*, 602-619.
- Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, *102*, 1134-1140.
- Lotto, A. J., Sullivan, S. C., & Holt, L. L. (2003). Central locus for non-speech effects on phonetic identification. *Journal of the Acoustical Society of America*, *113*, 53-56.
- Macmillan, N. A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (Ed.), *Categorical perception: The groundwork for cognition* (pp. 53-85). Cambridge, UK: Cambridge University Press.
- Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. *Psychological Review*, *84*, 452-471.
- Maddieson, I. (1984). *Patterns of sound*. Cambridge, UK: Cambridge University Press.
- Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception and Psychophysics*, *28*, 407-412.
- Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English 'l' and 'r'. *Cognition*, *24*, 169-196.
- Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *Journal of Neuroscience*, *3*, 1039-1057.
- Mehler, J., Jusczyk, P., Lambertz, C., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, *18*, 331-340.
- Mudry, K. M. (1978). *A comparative study of the response properties of higher auditory nuclei in anurans: Correlation with species-specific vocalizations*. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.
- Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, *Serinus canarius*. *Journal of Comparative Neurology*, *165*, 457-486.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, *13*, 253-260.
- Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, *61*, 1352-1361.
- Pisoni, D. B. (1987). General discussion of session 3: Dynamic aspects. In M. E. H. Schouten (Ed.), *Psychophysics of speech perception* (pp. 264-267). Dordrecht, Netherlands: Martinus Nijhoff.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*, 349-351.
- Remez, R. E. (1979). Adaptation of the category boundary between speech and nonspeech: A case against feature detectors. *Cognitive Psychology*, *11*, 38-57.
- Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *Journal of the Optical Society of America*, *43*, 495-501.

- Ryan, M. J., & Wilczynski, W. (1988). Coevolution of sender and receiver: Effect on local mate preference in cricket frogs. *Science*, *240*, 1786–1788.
- Sinex, D. G., & McDonald, L. P. (1988). Average discharge rate representation of voice-onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America*, *83*, 1817–1827.
- Sinex, D. G., & McDonald, L. P. (1989). Synchronized discharge rate representation of voice-onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America*, *85*, 1995–2004.
- Sinex, D. G., McDonald, L. P., & Mott, J. B. (1991). Neural correlates of nonmonotonic temporal acuity for voice onset time. *Journal of the Acoustical Society of America*, *90*, 2441–2449.
- Sinnot, J. M., & Brown, C. H. (1997). Perception of the English liquid /ra-la/ contrast by humans and monkeys. *Journal of the Acoustical Society of America*, *102*, 588–602.
- Smith, J. D., Kemler Nelson, D. G., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition*, *52*, 23–54.
- Steinschneider, M., Schroeder, C. E., Arezzo, J. C., & Vaughan, H. G., Jr. (1995). Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: Temporal response patterns. *Brain and Language*, *48*, 326–340.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.
- Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech* (pp. 1–38). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sussman, H. M. (1989). Neural coding of relational invariance in speech: Human language analogs to the barn owl. *Psychological Review*, *96*, 631–642.
- Sussman, H. M. (1994). The phonological reality of locus equations across manner class distinctions: Preliminary observations. *Phonetica*, *51*, 119–131.
- Sussman, H. M. (2002). Representation of phonological categories: A functional role for auditory columns. *Brain and Language*, *80*, 1–13.
- Sussman, H. M., Fruchter, D., & Cable, A. (1995). Locus equations derived from compensatory articulation. *Journal of the Acoustical Society of America*, *97*, 3112–3124.
- Sussman, H. M., Fruchter, D., Hilbert, J., & Sirosh, J. (1998). Linear correlates in the speech signal: The orderly output constraint. *Brain and Behavioral Sciences*, *21*, 241–299.
- Werner, L. A., & Bargones, J. Y. (1992). Psychoacoustic development of human infants. In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy research* (Vol. 7, pp. 103–145). Norwood, NJ: Ablex.
- Whalen, D. G., Abramson, A. S., Lisker, L., & Mody, M. (1993). *f* gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America*, *93*, 2152–2159.
- Winer, J. A. (1992). The functional architecture of the medial geniculate body and the primary auditory cortex. In D. B. Webster, A. N. Popper, & R. R. Fay (Eds.), *The mammalian auditory pathway. Neuroanatomy* (pp. 222–409). New York: Springer Verlag.
- Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America*, *60*, 1381–1389.